RNA folding: beyond Watson–Crick pairs

Eric Westhof* and Valérie Fritsch

Several crystal structures of RNA fragments, alone or in complex with a specific protein, have been recently solved. In addition, the structures of an artificial ribozyme, the leadzyme, and the cleavage product of a human pathogen ribozyme, have extended the structural diversity of ribozyme architectures. The attained set of folding rules and motifs expand the repertoire seen previously in tRNA structures.

Address: Institut de Biologie Moléculaire et Cellulaire du CNRS, UPR9002, 15, rue René Descartes, F-67084 Strasbourg-Cedex, France.

*Corresponding author.
E-mail: E.Westhof@ibmc.u-strasbg.fr

Structure 2000, **8**:R55–R65

0969-2126/00/$ – see front matter

Introduction

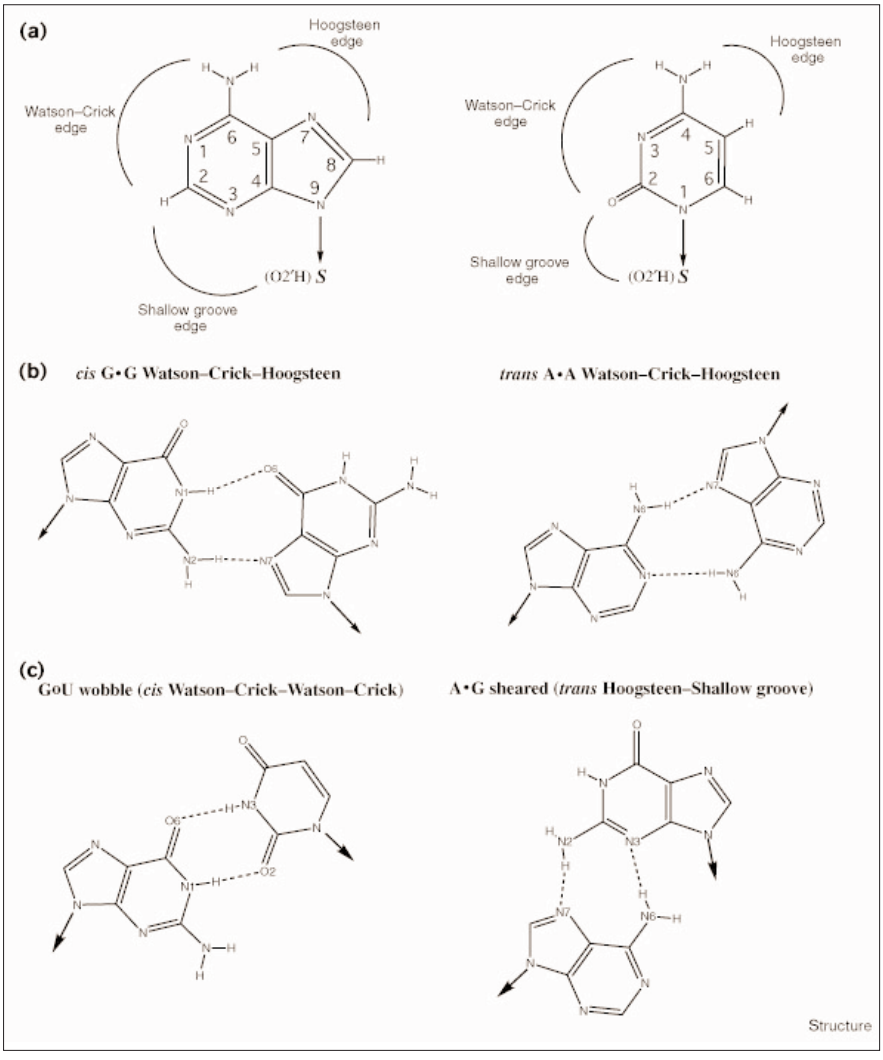
In the history of nucleic acid structural biology, surprises have been recurrent. Often discoveries were not the result of serendipity, but rather of crystallization experiments planned to observe other structures. Thus, the first crystal structure of a base pair in 1963, a major achievement in those days, revealed the existence of Hoogsteen pairing [1]. In the following years, several other non-Watson–Crick base pairs were observed in crystal structures [2]. Watson–Crick base pairs were first observed at atomic resolution in the crystal structures of ApU [3] and GpC [4]; the crystals contained a full sugar–phosphate backbone together with neutral bases. Following this accomplishment, structural biologists focused their attention on Watson–Crick pairs, consigning Hoogsteen pairs to the formation of triple helices [5] and in corners of tRNA structures [6]. It is now accepted that nucleic acid bases possess three edges onto which hydrogen bonding can occur (Figure 1): the Watson–Crick edge, the Hoogsteen edge and, finally, the shallow groove edge.

Improvements in RNA synthesis, purification techniques, X-ray sources, data collection and refinement have led to high-resolution RNA crystal structures with careful analysis of disorder [7–10]. The purpose of the present overview is not exhaustivity; instead we illustrate the recurrence of some structural aspects that have furthered our understanding of RNA architecture and folding since the early days of tRNA structural studies. Table 1 summarizes several new structural characteristics observed in recent crystal structures, some of which are described in the text and illustrated in the figures. The P4–P6 domain of group I introns and the hepatitis delta virus (HDV) structures have been carefully reviewed ([11,12]) as have the hammerhead ribozymes [13]. A survey of the crystal structures of helical domains is also available [14]. Some RNA motifs have been described and discussed by Moore [15] and all observed base pairs have been itemized in a review [16]. The use of non-Watson–Crick pairs as a recognition element in RNA–peptide/protein complexes has been discussed recently [17,18].

The flourishing diversity of base pairing

Because base pairing is so diverse, and because almost any combination of bases is observed in various geometries, some definitions are useful to characterize and organize base pairs. Firstly the hydrogen–bonding sites on the nucleic acid bases are distributed on three main edges (Figure 1a): the Watson–Crick edge, which presents the usual Watson–Crick sites; the Hoogsteen edge and the shallow groove edge. Pairs involving the shallow groove edge of one base, called sheared, are now frequently observed (Figure 1c). In such sheared pairs, one base presents its Watson–Crick sites to the deep groove and the other base exposes its Watson–Crick sites to the shallow groove. Purine–purine as well as pyrimidine–pyrimidine sheared pairs are observed [19,20]. Secondly, the bases can approach each other so that the sugars are on the same side or on opposite sides of a line median to the hydrogen bonds: in the first case, the pairing is called *cis* and in the second *trans* (Figure 1b). Accordingly, and for clarity, the Watson–Crick pairs will be noted A–U and G = C, the wobble pairs with a standard geometry GoU, and the non-Watson–Crick pairs by a • (Figures 1b,c). Finally, some pairs employ non-standard hydrogen-bonding rules (Figure 2). Thus, so-called bifurcated (or more appropriately 'chelated' [21]) hydrogen bonds have been observed; recently, in high-resolution structures of the loop E of rRNA and, previously, in the lower resolution studies of tRNAs (G18•Ψ55) [22]. The involvement of C–H bonds in some sort of hydrogen-bonding interaction cannot be dismissed owing to their frequent observation in high-resolution crystal structures [21] and to their surprising stability in long molecular dynamics simulations [23]. For example, the sheared G•A pair (Figure 1c) frequently exchanges with a sheared A•A pair in which the short distance between N7(A) and H–C2(A) indicates the presence of a C–H...N hydrogen bond. Disconcertingly, some pairs are mediated via one or more inserted water molecules. More puzzling still is a recent example of a Hoogsteen-like A•C base pair with no direct hydrogen bond but only water-mediated hydrogen bonds between N4(C) and N7(A) [19]. The example shown in Figure 3 is rather surprising and unexpected. Indeed, *cis* G•A pairs with two hydrogen bonds between the Watson–Crick

Figure 1



The three edges of bases and examples of *cis* and *trans* non-Watson–Crick pairs. **(a)** The three hydrogen-bonding edges of a purine base (left) or pyrimidine base (right). In the shallow groove edge, the ribose hydroxyl O2′H frequently participates in the hydrogen-bonded pair. **(b)** Examples of *cis* (left) and *trans* (right) base pairing. In *cis* pairing the glycosyl bonds are on the same side with respect to the hydrogen bonds linking the base pairs (or a median line between the two – rarely three – hydrogen bonds); in *trans*, the glycosyl bonds are on either side. The hydrogen-bonding sites, either Watson–Crick or Hoogsteen, together with the *cis* or *trans* orientation, are related to the local orientation of the strands. Thus, in the two examples shown, with the bases in the usual *anti* conformation with respect to the sugar, the strands are locally parallel. A parallel orientation of strands, instead of the usual antiparallel orientation, can occur locally following a reversal of the sugar–phosphate backbone (as in the eukaryotic loop E structure around the A•A pair) or globally because of the intricate folding of the strands (as for the invariant *trans* Watson–Crick R15•Y48 pair of tRNAs). **(c)** The standard wobble GoU (left) and sheared A•G (right) pairs.

sites of both bases and a distance between the C1′ atoms larger than those in Watson–Crick pairs (12.6 Å instead of 10.5 Å) occur in tRNA structures. In the water-mediated *cis* G•A pairs, however, there is only one direct hydrogen bond and the C1′...C1′ distance is even larger (14.8 Å) (Figure 3). The opening occurs in the shallow groove side, a fact that might be related to the presence of magnesium ions bound in the deep groove. In addition, the *cis* water-mediated G•A pair is sandwiched between two bifurcated pairs. Formation of triples in the deep groove has been well documented since the early work of Arnott [24] and the tRNA structures [6]. However, it now appears that triples in the deep groove exploit C–H...O hydrogen bonds (e.g., C1072•C1092 in Figure 4); in the P4–P6 structure [25], a homologous U259•U107 base pair occurs in which the N3...H–N4 hydrogen bond has been replaced by N3–H...O4 while the C5–H...O2 hydrogen bond is maintained.

Hydrogen bonding in the shallow groove is common and versatile

Among the important findings from the recent RNA structures was the elucidation of the subtle and unforeseen roles of hydrogen bonding in the shallow groove of RNA helices. These roles were recently discussed extensively [16] and we present here a rapid overview. Hydrogen bonds in the shallow groove all involve the O2′ hydroxyl group, and adenine residues are the most frequent bases found to interact with the shallow groove edge of another base (for side-by-side pairs of the 'AA-platform' motif, see below) [26]. The adenine base interacts via its Watson–Crick sites (N1 and N6), Hoogsteen sites (N6 and N7) or shallow groove sites (N3 and C2–H) with the shallow groove sites (N3(R), N2(G), O2(Y) and the O2′ hydroxyl) of another base, which is often itself engaged in a Watson–Crick or Hoogsteen pair. The type of atom interacting with the hydroxyl group is different in

Table 1

Crystal structures of RNA with a short description of some of the new structural aspects they display.

Name	Size (nt)	Resolution (Å)	Motifs
Hammerhead ribozyme*	59	2.6	GA tandem of sheared G•A Single hydrogen bond U•A GAAA tetraloop GAAA/G = C pairs contact
P4–P6 domain†	259	2.8	Ribose zipper AA platform GAAA/11-nucleotide motif Adenine-rich bulge + Mg^{2+} ions P5abc three-way junction J5/5a bent internal loop
Hepatitis delta virus ribozyme‡	72	2.3	Double pseudo-knot Single-stranded adenines in the shallow groove
Leadzyme ribozyme§	24	2.7	Three-base-pair parallel helix Single hydrogen bond C•A Pb^{2+}-binding site
Beet Western Yellow virus pseudo-knot#	28	1.6	Single-stranded adenines in the shallow groove Quadruple pair
5S rRNA Loop E¶	62 22	3.0 1.5	Cross-strand purine stacks Bifurcated G•U and G•G Water-mediated G•A Mg^{2+}-binding sites in deep groove
Sarcin/ricin loop¥	29 (rat) 27 (*E. coli*)	2.1 1.11	S turn (S motif) No direct hydrogen bond A•C Triple between UG platform and a Hoogsteen A
HIV TAR RNA**	27	1.3	Ca^{2+}-binding sites Co-axial and continuous helix despite the asymmetric three-nucleotide bulge
HIV DIS††	22	2.3	Bulge A 5′ to open G•A Mg^{2+}-binding sites between N7 of G and proR oxygen of bulging A
SRP Helix 6‡‡	29	2.0	Tandem of open G•A between bifurcated G•G Tandem of protonated CoA^{+} Neutral C•A (N4(C)...N1(A))

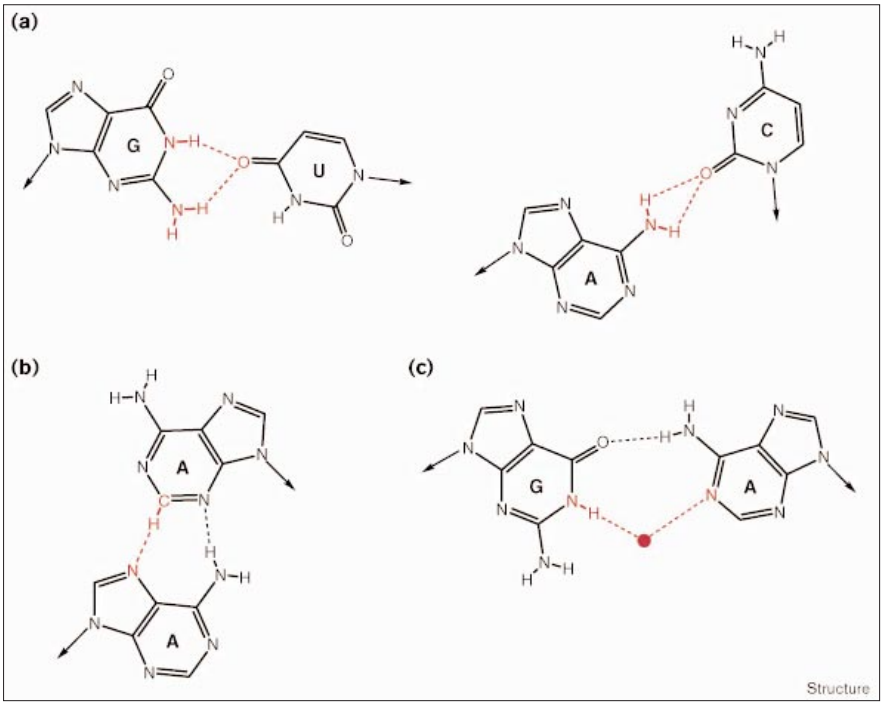
References: *[33,59]; †[25,26,42,43]; ‡[54]; §[60]; #[61]; ¶[35]; ¥[19,20]; **[62]; ††[63]; ‡‡[64].

the *cis* or *trans* pairs. Thus, in a *cis* Watson–Crick–shallow groove pair, the N1 nitrogen of adenine binds the hydroxyl group, whereas in a *trans* pair it is the N6 amino group that participates (Figure 5a). Similarly, in a *trans* shallow groove–shallow groove pair, the N1 atom binds the hydroxyl group, but in a *cis* pair it is the N3 atom (Figure 5b). With the bases in the *anti* conformation, these choices are related to the relative orientations of the paired strands (Figure 6). The sheared G•A pairs that close GNRA tetraloops belong to the family of *trans* Hoogsteen–shallow groove pairs. The surprising AA-platform motif [26], in which two consecutive nucleotides stay side-by-side in the same plane, involves a *cis* Hoogsteen–shallow groove contact between the 3′-base and the 5′-base. As in other sheared base pairs, the 3′-base of the AA platform is engaged in a Watson–Crick or Hoogsteen pairing. In fact, side-by-side platforms are not restricted to 5′-AA-3′ dinucleotides; 5′-GU-3′ platforms are observed in the sarcin loop [19] and in a complex formed between 23S rRNA and the ribosomal protein L11 [27] (Figure 7).

Interstrand or cross-strand stacking

In standard B-DNA structures, base stacking occurs mainly between bases on the same strand with the sequence having only a minor influence (intrastrand stacking). In RNA helices (or A-DNA helices), however, base stacking is strongly influenced by sequence: generally, in 5′-R–Y-3′ steps one observes intrastrand stacking, and in 5′-Y–R-3′ steps there is definite interstrand stacking. This tendency is accentuated in non-Watson–Crick pairs. A well-described example is that of wobble GoU pairs, for which there is pronounced interstrand stacking between the guanine residues in tandem GoU pairs with the sequence order 5′-UG-3′ [28]. Stretches of non-Watson–Crick pairs

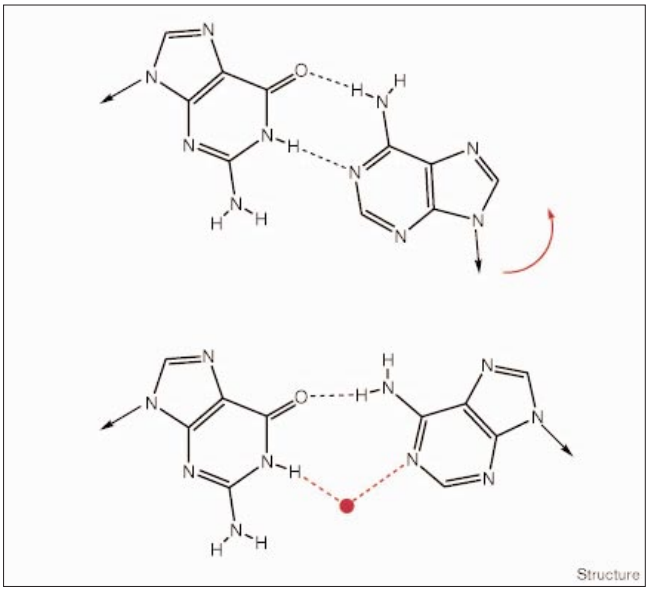
Figure 2



Non-Watson–Crick base pairs sometimes employ non-standard hydrogen bonds. **(a)** Chelated (or bifurcated) hydrogen bonds, where two hydrogen atoms point at an angle to a single acceptor oxygen atom. **(b)** The controversial C–H...N/O hydrogen bond. These are quite frequently observed in crystal structures and have been shown to be stable in molecular dynamics simulations. **(c)** A single direct hydrogen bond between two bases, with a second bond occurring via an inserted water molecule.

display pronounced purine stacks; one of the best examples is loop E of 5S rRNAs [19]. The stabilizing effect of several layers of purine stacks is also seen in bent junctions. For example, in tRNAs the two bulging residues 59 and 60 stack on each other and on the two non-Watson–Crick *trans* pairs U8•A14/R15•Y48, stabilising the 90° interface between the two arms. In the recent nuclear magnetic resonance (NMR) structure of a complex between protein L30 and its regulatory RNA, the 130° bend between the two helical regions is stabilized by an interesting three-layer purine stack [29].

Figure 3



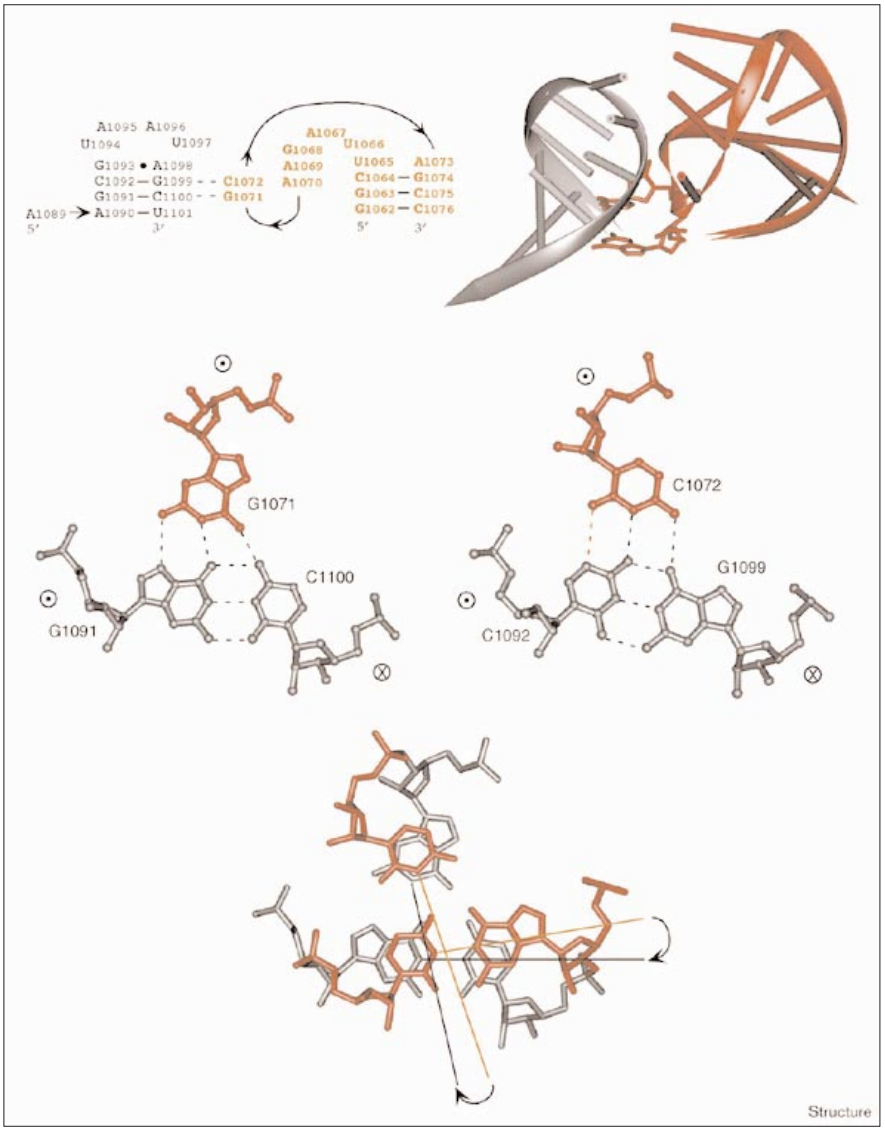
The opening rotation movement that transforms a *cis* Watson–Crick G•A pair into an opened and water-mediated G•A pair. The water molecule and associated hydrogen bonds are shown in red.

RNA–RNA recognition motifs

Whereas the Watson–Crick pairs between complementary bases are a necessity for forming the helical framework of a complex RNA, the non-Watson–Crick pairs are pivotal in RNA–RNA and RNA–protein recognition. The complementary Watson–Crick base pairs, with *cis* glycosyl bonds, form the only set of pairs that are isosteric in antiparallel helices. Thus, they promote the formation of helices with quasi-regular sugar–phosphate backbones that define the secondary structure. In single-stranded RNA molecules, stacking and base pairing drive the folding of the chain on itself through the formation of helical regions linked by non-helical elements, hairpin loops, internal bulges and multiple junctions. RNA tertiary structure will therefore comprise RNA–RNA interactions involving either two helices, two unpaired regions, or one unpaired region and a double-stranded helix [11,30].

The interactions between two helices are basically of two types: either two helices with a contiguous strand stack on each other or two distant helices position themselves so

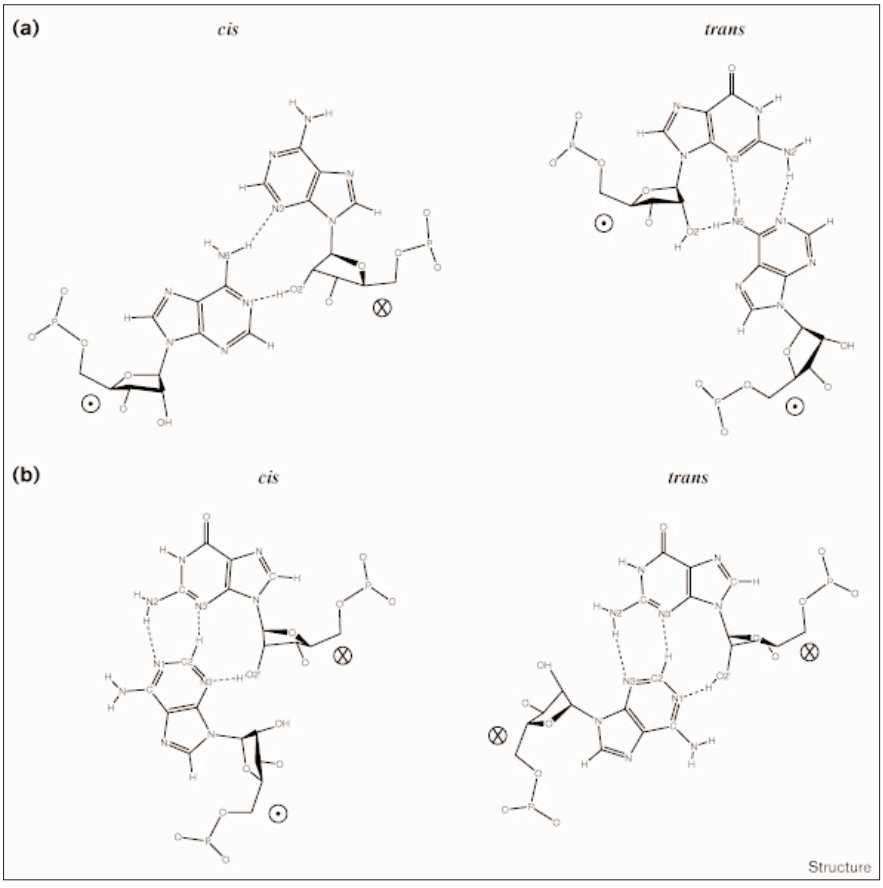
Figure 4



The bent hairpin loop (in red) present in the complex between protein L11 and its binding region in the 23S rRNA. Because of the 180° bend of the nine-nucleotide-loop, residues G1071 and C1072 are each able to form a triple interaction with a G = C pair stacked on a C = G pair. Notice also that the rotation between the two single-stranded residues follows that between the two stacked pairs (lower panel). The six-nucleotide-loop (in black) is also of interest: it is closed by a sheared G•A pair and the other four residues adopt the same conformation as the anticodon loop of tRNAs between residues U33 and 36 (i.e., U1094 forms a U-turn; U1094 and U1097 presents a O...H–C contact between the O2(U1094) and C6(1097)). The same loop with a similar conformation is found in U2 snRNA [65] and in loop 715 of the 23S rRNA [66]. Very similar loops are also found in aptamers against aminoglycosides [67]. The symbols next to the sugars indicate the strand direction: a dot shows that the 5′ to 3′ direction points towards the reader; a cross indicates that the 5′ to 3′ direction points away from the reader into the page.

that hydrogen bonding between the sugar–phosphate backbones occurs in the shallow grooves. The second type of contact, observed as intermolecular crystal contacts [14], is beautifully illustrated intramolecularly in the P4–P6 structure [25]. An unpaired region belongs to either a single-stranded stretch (forming an internal loop or a bulge) or a hairpin loop closing a helix. Interactions between two unpaired regions can be mediated by standard Watson–Crick pairing leading to the formation of the tertiary motif called a pseudo-knot, if a single loop is involved [31] (with the possibility of co-axial stacking between the helices), or alternatively to loop–loop motifs. Interactions between an unpaired region and a double-stranded helix can lead to various types of motifs, and always involve non-Watson–Crick pairs. Hydrogen bonding of a single-stranded stretch, to sites either in the deep or the shallow groove of a double helix, leads to the formation of triples. Since the determination of the tRNA structure, those formed in the deep groove are well known [6] (Figure 4). As discussed above, the recent structures display a rich variety of triples formed in the shallow groove (Figures 5,6). Two RNA–RNA self-assembly motifs are known in which the unpaired region constitutes a terminal hairpin loop: GNRA tetraloops bind the shallow groove of an RNA helix [32,33], whereas GAAA tetraloops bind to a specific 11-nucleotide receptor [25,34]. Both motifs had been predicted on the basis of sequence analysis, coupled to molecular modelling, chemical probing, and *in vitro* selection studies [32,34]. The GNRA motif was first observed as an intermolecular contact in crystals of the hammerhead ribozyme [33], whereas the GAAA motif (Figure 8) links the two main helical domains of the P4–P6

Figure 5



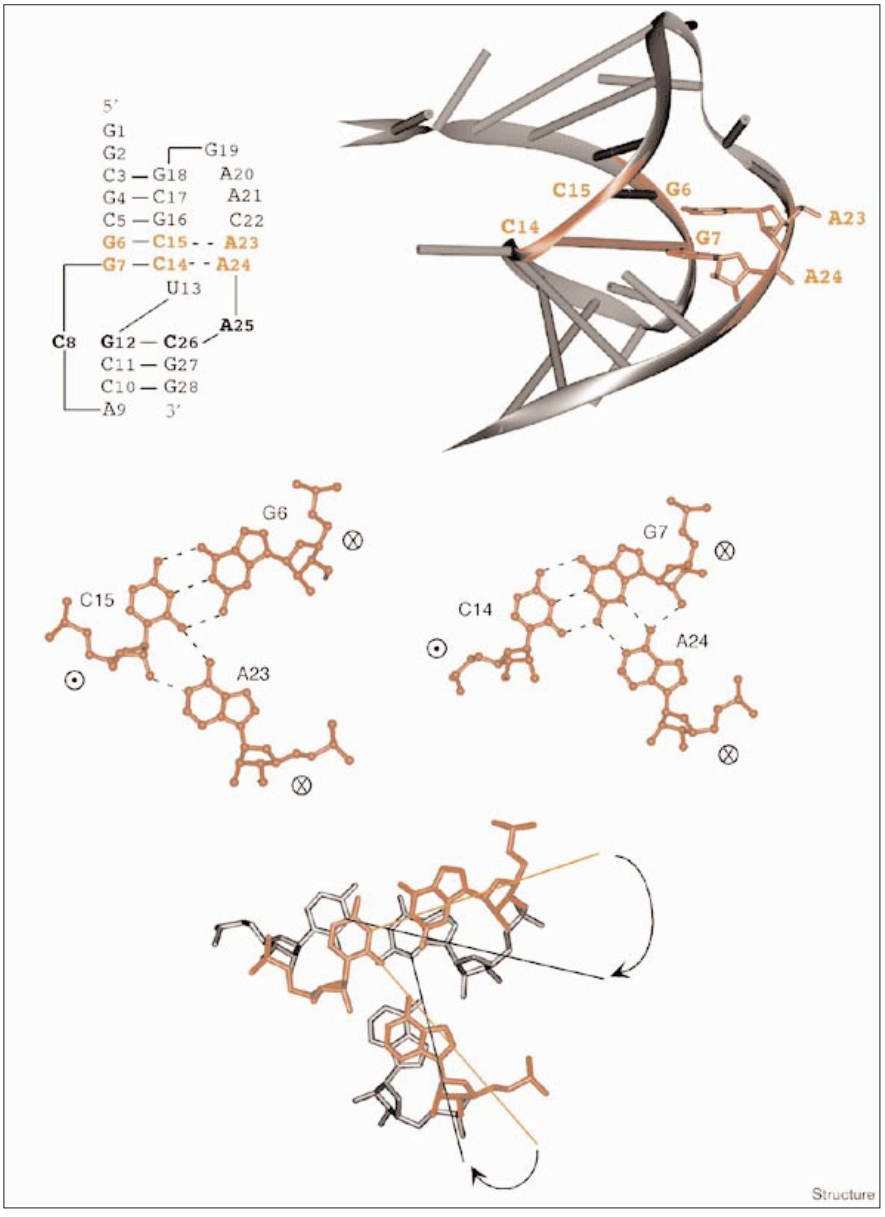
Idealized drawings of shallow groove pairs. **(a)** Watson–Crick–shallow groove pairs and **(b)** shallow groove–shallow groove pairs, both in *cis* and *trans*. The strand direction is indicated as described in Figure 4. Note the systematic use of the hydroxyl group O2′-H. For more information and details see [16].

structure [25]. Internal loops also form three-dimensional motifs, as their bases are engaged in non-Watson–Crick pairings leading to compact and helix-like regions which often bind magnesium ions. A well-described and analysed system is loop E of 5S rRNA [35,36]. Likewise, the P4–P6 structure contains an adenine-rich loop that is organized around two magnesium ions and presents adenine residues for interacting with a helix [25].

Unpaired regions of the secondary structure are structured

In RNA, secondary structure is usually defined in terms of contiguous regions of *cis* Watson–Crick base pairs (including wobble GoU pairs) forming helices. Formally, the RNA folding problem is simpler than the protein folding problem [37,38]. Indeed, the energy content of the secondary structure is large compared with that of the tertiary structure, therefore, the energy of the interactions maintaining the three-dimensional architecture can be considered as a perturbation on the energy of the overall system. Experimentally, this hierarchical view of RNA folding is observed in UV melting of folded RNAs, where the cooperative melting of the tertiary structure is observed first before the broad and sequential melting of the secondary structure elements [39–41]. The melting of the tertiary structure also depends strongly on divalent ion concentrations, especially magnesium ions, implying that specific ion-binding sites are created during tertiary folding [37,40,42–44]. On the other hand, monovalent ions influence the stability of secondary structure elements [37,38,45]. The distinction between two-dimensional and three-dimensional structures is commonly observed during *in vitro* experiments (for a recent appraisal, see [46]). The hierarchy in RNA folding forms the basis of a modelling approach in which preformed RNA modules are assembled into complex architectures via defined tertiary contacts [30,32]. It is now clear that the single-stranded interhelical segments are rarely unpaired and instead form structured regions that tend to be helical-like and bind magnesium ions. This organization is beautifully illustrated by the loop E domain of 5S rRNA. Solution data originally concluded that magnesium ions were necessary for the structuring of this loop [47,48]. Later, NMR evidence indicated the presence of several non-Watson–Crick pairs in the loop E of eukaryotic 5S rRNA [49] and their presence was subsequently confirmed by high-resolution X-ray crystallography of the sarcin loop [19]. Thus, one can expect that most of the unpaired regions in the secondary structures of ribosomal RNAs or large catalytic RNAs are

Figure 6



The high-resolution crystal structure of a pseudo-knot. Although the general features of the architecture follow the modelled structures, the high-resolution structure reveals some unexpected facts. Firstly, the secondary structure is not fully obeyed: a Watson–Crick base pair between U13 and A25 was expected, and A9, instead of being stacked below the last base pair (C10–G28), would have been expected to stretch the deep groove of the three-base-pair helix. Secondly, among the residues stretching the shallow groove of the five-base-pair helix, two adenines make precise interactions with two consecutive G = C pairs. For both of them, the interactions involve the Watson–Crick sites of the adenine: the N6 with O2(Y) or N3(R) as well as the N1 with the ribose hydroxyl O2′ or the guanine amino group. Interestingly, the two single-stranded adenines follow the right-handed rotation between the two base pairs (lower panel). Thus, while A23, in *cis* with respect to C15, is antiparallel to it, A24, in *trans* with respect to G7, is parallel to it. A very similar type of pattern occurs in the tertiary motif between GYRA and helical base pairs. The strand direction is indicated as described in Figure 4.

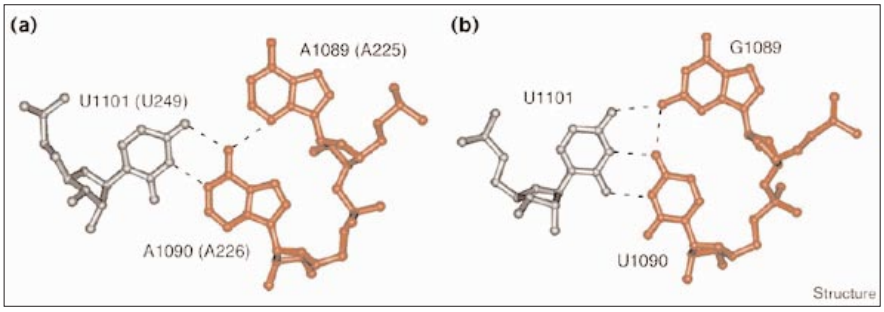
in fact structured and organized. It was suggested that the loop E motif, or the S motif, is an organizing motif of the structure of multihelix loops in 16S and 23S rRNAs [50]. In the recent crystal structure of the 70S ribosome [51], one such motif was indeed seen in the electron density.

Conclusions

Until very recently, the only known RNA structures were those of the RNA building blocks, some RNA fragments, and tRNA. The stereochemical rules, established several years ago on the basis of X-ray diffraction studies of nucleosides and nucleotides [52,53], are confirmed by the recent structures (see Figure 9). The bases adopt overwhelmingly the *anti* conformation with respect to the sugar. Although bases in the *syn* conformation are more frequent in NMR-derived structures than X-ray structures, their frequency is still very low and their presence rather exceptional. The sugar puckers are commonly in the C3′-*endo* conformation, even outside helical domains, and the C2′-*endo* pucker is restricted to tight turns or loops. The other torsional angles also adopt restricted conformations, with the torsion angles about the C3′–O3′ and C5′–O5′ bonds mainly in the *trans* region.

At the tertiary structure level, the tRNAs revealed several folding rules that are still valid. Firstly, neighbouring

Figure 7

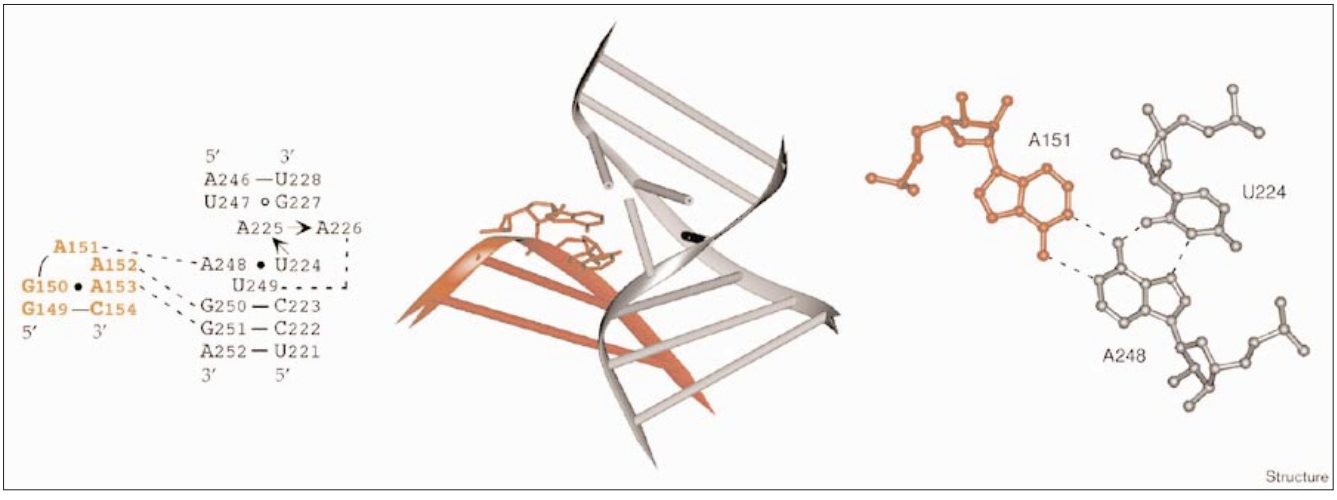


AA platforms are not restricted to AA dinucleotides. **(a)** A typical AA platform. **(b)** A 5′-GU-3′ side-by-side contact (the arrangement was reconstructed as the coordinates are not available for this structure). Notice how the 3′-base forms a *cis* Watson–Crick pair while the 5′-base is poised to interact with its Watson–Crick sites. In the recent crystal structure of the complex between threonyl-tRNA synthetase and its cognate tRNA [68], the major identity elements in the anticodon loop, G35 and U36, form a platform with a N2(G35)...O4(U36) hydrogen bond which stacks on a flat surface of conserved hydrophobic residues. The Watson–Crick sites, N1(G35) and O2(U36), make specific interactions with conserved residues, Glu600 and Arg609, respectively.

helices in the secondary structure, with no unpaired residues on one strand, stack in a coaxial manner (e.g., as seen in helices 2 and 4 of the HDV ribozyme [54]). Secondly, non-Watson–Crick base pairs are used systematically for tertiary contacts between and within domains. Thus, hairpin loops frequently contain non-Watson–Crick base pairs at the interface between the helix and loop region (e.g., the *trans* Hoogsteen pair in the thymine loop, the bifurcated G•U pair in the UNCG tetraloop, or the 32...38 base pair in the anticodon loop [55]). At the junction between stacked helices, non-Watson–Crick pairs modulate the stacking and help formation of further tertiary contacts like base triples. For example, in the structures of tRNA, at the junction between the dihydrouridine and anticodon loop, a *cis* Watson–Crick like G•A pair (Figure 3) is frequently observed. Likewise, in the L11 complex, between the stacked helices B and C, there is a sheared G•A pair (Figure 1) [56]. Thirdly, loop–loop interactions rely on Watson–Crick pairs (e.g., G19–C56) as well as on more complex interactions like intercalation or non-Watson–Crick pairs (G18•Ψ55). Similarly, in the HDV ribozyme [54] there is a two-base-pair pseudo-knot mediated via two invariant Watson–Crick G = C pairs between a loop and a single-stranded region. Fourthly, the U-turn is a highly recurrent element of RNA structure and is of great importance in folding (e.g., two occurrences in the tRNA structure and almost one occurence in any large structure). Fifthly, the 2′-hydroxyl group is systematically and astutely used in turns and close tertiary approach. Finally, C2′-*endo* sugar puckers appear in loops and non-helical regions of the sugar–phosphate backbone.

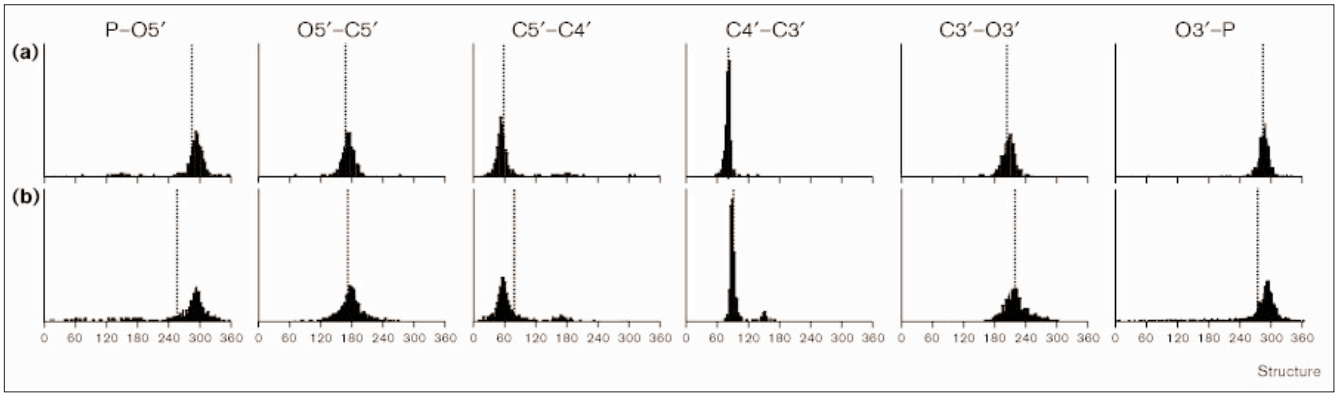
The new crystal structures, however, have disclosed new vistas on RNA stereochemistry and introduced new

Figure 8



The specific receptor for the GAAA tetraloop the 11-nucleotide motif, as seen in the crystal structure of the P4–P6 domain. The AA platform (A225–A226) is sandwiched between a *trans* Hoogsteen pair (A248•U224) and a wobble pair (U247∘G227). A248 forms a *trans* Watson–Crick A•A pair with A151, the apical base of the GAAA tetraloop. The bulging U249 base forms a *cis* Watson–Crick pair with A226 (see Figure 7).

Figure 9



Distribution of angle values for each torsion angle in the sugar–phosphate backbone calculated in the high-resolution crystal structures of RNAs present in the Nucleic Acid Database. **(a)** The torsion angles from the structures of RNA helical fragments. The plot encompasses about 630 nucleotides and includes some non-Watson–Crick pairs. **(b)** The torsion angles for the large RNA structures (tRNAs, P4–P6, ribozymes etc.) are plotted. These encompass about 1300 nucleotides and include several non-helical residues. The dotted lines indicate the mean value of each torsion angle. It is apparent that the torsion angles about the P–O bonds are centred on −60°, those about the C–O bonds on 180°, and those about the C–C bonds around +60° (a value for C4′–C3′ of +80° corresponds roughly to a sugar pucker in the C3′-*endo* domain with a standard amplitude of pucker [69]). This was first noticed in 1969 by Sundaralingam on the very meagre set of structures available then [52] and formed the basis of the 'rigid' nucleotide concept [53]. In non-helical regions, other alternative torsion angles are observed but their proportions are still rather low.

unexpected rules of RNA folding. Firstly, some recognition and folding motifs are clearly used recurrently. In addition, like Russian dolls, three-dimensional motifs are assembled using frequent smaller stereochemical motifs. Thus, GNRA tetraloops exploit the celebrated U-turn motif. Similarly, a sheared G•A pair does not fit alone within a helix, because of the short distance between the O3′ atom of the guanine and the 5′-phosphate of the adenine [57]. Consequently, sheared G•A pairs occur in tandem (5′-GA-3′) or in conjunction with another non-Watson–Crick pair, such as a *trans* Hoogsteen base pair, 3′ to the guanine. The question remains open as to the number of recurrent motifs still to be discovered [15]. Secondly, several examples exist of intermolecular contacts that organize the crystalline packing, which are also used intramolecularly in a supramolecular fashion. A prominent example is the ribose zipper motif [25] for maintaining parallel stacking of helices; another example is given by members of the GNRA tetraloop family which bind to their respective receptors intramolecularly or intermolecularly. Thirdly, in the RNA structural world, the appearance of non-Watson–Crick pairs, in internal loops, junctions or hairpin loops, is now so prevalent that the words 'mismatch' and 'mispair' have lost their meaning. It now seems that one of the main functional and structural roles of RNA helices is to subtend a three-dimensional scaffold, critically maintained by non-Watson–Crick pairs, presenting recognition and binding motifs made of non-Watson–Crick pairs. In this respect, the pairs formed in the shallow groove, like the sheared pairs or the side-by-side bases in the AA-platform motif, are especially important as their Watson–Crick sites are available for RNA or protein binding. In these shallow groove pairs, except in the AA-platform motif, the O2′ hydroxyl group is implicated, emphasizing how functionally specific to RNA these motifs are. Interestingly, the only theoretical prediction of the side-by-side bases, an arrangement that does not use the hydroxyl group, was proposed for some DNA-specific sequences [58].

Presently, it is not clear how the appropriate functional equilibrium between Watson–Crick and non-Watson–Crick pairs is determined during evolution (natural or artificial). Divalent ions could be an important factor for the maintenance of the functional distribution of base pairs. Indeed, the recent crystal structures provide a wealth of new structural information and insight into divalent ion binding to RNA, especially for key magnesium ions. Although magnesium ions bind frequently to (and often link) the anionic phosphate oxygen atoms and the Hoogsteen sites of guanines, the non-Watson–Crick pairs, because of their effects on groove size and the sugar–phosphate backbone path, often mould ion-binding cavities. Thus, the balance between Watson–Crick and non-Watson–Crick pairs is strongly dependent on the concentration of divalent ions.

Nowadays, the structural ease with which RNA forms non-Watson–Crick pairs and the variety of these pairs is still a curse for RNA crystallographers, who often observe extended duplexes with several non-Watson–Crick pairs instead of the hoped for monomeric hairpin fold or motif. Thus, in continuation with its history, RNA crystallography

persists in surprising us, keeping us away from despondency by the excitement and fascination of unexpected RNA structures.

Acknowledgements
We wish to thank Pascal Auffinger for Figure 9 and Neocles Leontis for his constant input on the classification of non-Watson–Crick pairs. EW wishes to thank the Institut Universitaire de France for support.

References

1. Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta crystallogr.* **16**, 907-916.
2. Voet, D. & Rich, A. (1970). The crystal structures of purines, pyrimidines and their intermolecular complexes *Prog. Nucleic Acid Res. Mol. Biol.* **10**, 183-265.
3. Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Kim, J.J. & Rich, A. (1976). RNA double-helical fragments at atomic resolution. I. The crystal and molecular structure of sodium adenylyl-3′,5′-uridine hexahydrate *J. Mol. Biol.* **104**, 109-144.
4. Rosenberg, J.M., Seeman, N.C., Day, R.O. & Rich, A. (1976). RNA double-helical fragments at atomic resolution. II. The crystal structure of sodium guanylyl-3′,5′-cytidine nonahydrate *J. Mol. Biol.* **104**, 145-167.
5. Sun, J.S., Garestier, T. & Hélène, C. (1996). Oligonucleotide-directed triple-helix formation *Curr. Opin. Struct. Biol.* **6**, 327-333.
6. Saenger, W. (1988). *Principles of Nucleic Acid Structure*, 2nd edn. Springer-Verlag, New-York.
7. Klosterman, P.S., Shah, S.A. & Steitz, T.A. (1999). Crystal structures of two plasmid copy control related RNA duplexes: an 18 base pair duplex at 1.20 Å resolution and a 19 base pair duplex at 1.55 Å resolution. *Biochemistry* **38**, 14784-14792.
8. Shah, S.A. & Brünger, A.T. (1999). The 1.8 Å crystal structure of a statically disordered 17 base-pair RNA duplex: principles of RNA crystal packing and its effect on nucleic acid structure. *J. Mol. Biol.* **285**, 1577-1588.
9. Masquida, B., Sauter, C. & Westhof, E. (1999). A sulfate pocket formed by three G◦U pairs in the 0.97 Å resolution X-ray of a nonameric RNA. *RNA* **5**, 99-112.
10. Mueller, U., Schübel, H., Sprinzl, M. & Heinemann, U. (1999). Crystal structure of acceptor stem of $tRNA^{Ala}$ from *Escherichia coli* shows unique G◦U wobble base pair at 1.16 Å resolution *RNA* **5**, 670-677.
11. Batey, R.T., Rambo, R.P. & Doudna, J.A. (1999). Tertiary motifs in RNA structure and folding. *Angew Chem. Int. Ed. Engl.* **38**, 2326-2343.
12. Hermann, T. & Patel, D. (1999). Stitching together RNA tertiary architectures. *J. Mol. Biol.* **294**, 829-849.
13. Wedekind, J.E. & McKay, D.B. (1998). Crystallographic structures of the hammerhead ribozyme: relationship to ribozyme folding and catalysis. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 475-502.
14. Masquida, B. & Westhof, E. (1999). Crystallographic structures of RNA oligonucleotides and ribozymes. In *Oxford Handbook of Nucleic Acid Structures.* (Neidle, S., ed.), pp. 533-565, Oxford University Press, Oxford, UK.
15. Moore, P.B. (1999). Structural motifs in RNA. *Annu. Rev. Biochem.* **67**, 287-300.
16. Leontis, N.B. & Westhof, E. (1998). Conserved geometrical base pairing patterns in RNA. *Quart. Rev. Biophysics* **31**, 399-455.
17. Hermann, T. & Westhof, E. (1999). Non-Watson–Crick base pairs in RNA–protein recognition *Chem. Biol.* **6**, R335-R343.
18. Hermann, T. & Patel, D. (2000). Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820-825.
19. Correll, C.C., Wool, I.G. & Munishkin, A. (1999). The two faces of the *Escherichia coli* 23S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275-287.
20. Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., Wool, I.G. & Steitz, T.A. (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors *Proc. Natl Acad. Sci. USA* **95**, 13436-13441.
21. Jeffrey, G.A. (1988). *An Introduction to Hydrogen Bonding*, 2nd edn. Oxford University Press, Oxford, UK.
22. Quigley, G.J. & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science* **194**, 796-806.
23. Auffinger, P., Louise-May, S. & Westhof, E. (1996). Hydration of C-H groups in tRNA. *Faraday Discuss.* **103**, 151-173.
24. Arnott, S., Bond, P.J., Selsing, E. & Smith, P.J. (1976). Models of triple-stranded polynucleotides with optimised stereochemistry *Nucleic Acids Res.* **3**, 2459-2470.
25. Cate, J.H., *et al.*, & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain – principles of RNA packing. *Science* **273**, 1678-1685.
26. Cate, J.H., *et al.*, & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696-1699.
27. Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11–RNA complex. *Cell* **97**, 491-502.
28. Masquida, B. & Westhof, E. (2000). On the wobble G◦U and related pairs. *RNA* **6**, 9-15.
29. Mao, H., White, S.A. & Williamson, J.R. (1999). A novel loop–loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat. Struct. Biol.* **6**, 1139-1147.
30. Westhof, E. & Michel, F. (1994). Prediction and experimental investigation of RNA secondary and tertiary foldings. In *RNA–Protein Interactions.* (Nagai, K. & Mattaj, I.W., eds), pp. 25-51, IRL Press, Oxford.
31. Hilbers, C.W., Michiels, P.J. & Heus, H.A. (1998). New developments in structure determination of pseudoknots *Biopolymers* **48**, 137-153.
32. Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585-610.
33. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Model for a tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111-113.
34. Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276-1285.
35. Correll, C.C., Freeborn, B., Moore, P.B. & Steitz, T.A. (1997). Metals, motifs and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705-712.
36. Leontis, N.B. & Westhof, E. (1998). The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134-1153.
37. Brion, P., Michel, F., Schroeder, R. & Westhof, E. (1999). Analysis of the cooperative thermal unfolding of the td intron of bacteriophage T4 *Nucleic Acids Res.* **27**, 2494-2502.
38. Tinoco, I., Jr. & Bustamante, C. (1999). How RNA folds. *J. Mol. Biol.* **293**, 271-281.
39. Jaeger, L., Westhof, E. & Michel, F. (1993). Monitoring of the cooperative unfolding of the sunY group I intron of the bacteriophage T4. *J. Mol. Biol.* **234**, 331-346.
40. Tanner, M. & Cech, T. (1996). Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium *Azoarcus*. *RNA* **2**, 74-83.
41. Banerjee, A.R., Jaeger, J.A. & Turner, D.H. (1993). Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. *Biochemistry* **32**, 153-163.
42. Cate, J.H. & Doudna, J.A. (1996). Metal-binding sites in the major groove of a large ribozyme domain. *Structure* **4**, 1221-1229.
43. Cate, J.H., Hanna, R.L. & Doudna, J.A. (1997). A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.* **4**, 553-558.
44. Misra, V.K. & Draper, D.E. (1998). On the role of magnesium ions in RNA stability. *Biopolymers* **48**, 113-135.
45. Costa, M., Fontaine, J.M., Loiseaux-de Goer, S. & Michel, F. (1997). A group II self-splicing intron from the brown alga *Pylaiella littoralis* is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J. Mol. Biol.* **274**, 353-364.
46. Silverman, S.K., Zheng, M., Wu, M., Tinoco, I.Jr. & Cech, T.R. (1999). Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. *RNA* **5**, 1665-1674.
47. Leontis, N.B., Ghosh, P. & Moore, P.B. (1986). Effect of magnesium ion on the structure of the 5S RNA from *Escherichia coli*. An imino proton magnetic resonance study of the helix I, IV, and V regions of the molecule. *Biochemistry* **25**, 7386-7392.
48. Romby, P., *et al.*, & Ehresmann, B. (1988). Higher order structure of chloroplastic 5S ribosomal RNA from spinach. *Biochemistry* **27**, 4721-4730.
49. Wimberly, B., Varani, G. & Tinoco, I., Jr. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078-1087.
50. Leontis, N.B. & Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* **283**, 571-583.
51. Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N. & Noller, H.F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095-2104.

52. Sundaralingam, M. (1969). Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside, mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers* **7**, 821-860.
53. Sundaralingam, M. (1973). The concept of a conformationally 'rigid' nucleotide and its significance in polynucleotide conformational analysis. In *Conformation of Biological Molecules and Polymers*. (Bergmann, V.E.D. & Pullmann, B. eds), pp. 417-456, The Israel Academy of Sciences and Humanities, Jerusalem.
54. Ferré-D'Amaré, A.R., Zhou, K.H. & Doudna, J.A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-574.
55. Auffinger, P. & Westhof, E. (1999). Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.* **292**, 467-483.
56. Conn, G.L., Draper, D.E., Lattman, E.E. & Gittis, A.G. (1999). Crystal structure of a conserved ribosomal protein–RNA complex. *Science* **284**, 1171-1174.
57. Gautheret, D., Konings, D. & Gutell, R.R. (1994). A major family of motifs involving G•A mismatches in ribosomal RNA. *J. Mol. Biol.* **242**, 1-8.
58. Kuryavyi, V.V. & Jovin, T.M. (1995). Triad-DNA: a model for trinucleotide repeats. *Nat. Genet.* **9**, 339-341.
59. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991-1002.
60. Wedekind, J.E. & McKay, D.B. (1999). Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis *Nat. Struct. Biol.* **6**, 261-268.
61. Su, L., Chen, L., Berger, J.M. & Rich, A. (1999). Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol.* **6**, 285-292.
62. Ippolito, J.A. & Steitz, T.A. (1998). A 1.3 Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation *Proc. Natl Acad. Sci. USA* **95**, 9819-9824.
63. Ennifar, E., *et al*., & Dumas, P. (1999). The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure* **7**, 1439-1449.
64. Wild, K., Weichenrieder, O., Leonard, G.A. & Cusack, S. (1999). The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Structure* **7**, 1345-1352.
65. Stallings, S.C. & Moore, P.B. (1997). The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA. *Structure* **5**, 1173-1185.
66. Culver, G.M., Cate, J.H., Yusupova, G.Z., Yusupov, M.M. & Noller, H.F. (1999). Identification of an RNA–protein bridge spanning the ribosomal subunit interface. *Science* **285**, 2133-2136.
67. Walter, F., Vicens, Q. & Westhof, E. (1999). Aminoglycoside–RNA interactions. *Curr. Opin. Chem. Biol.* **3**, 694-704.
68. Sankaranarayanan, R., *et al*., & Moras, D. (1999). The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97**, 371-381.
69. Westhof, E. & Sundaralingam, M. (1980). Interrelationships between the pseudorotation parameters and the geometry of the furanose ring. *J. Am. Chem. Soc.* **102**, 1493-1500.